CHALMERS
UNIVERSITY OF TECHNOLOGY

Content placement in 5G-enabled edge/core datacenter networks resilient to link cut attacks

Received: 17 January 2020 | Accepted: 20 January 2020 | Published on: 4 February 2020

DOI: 10.1002/net.21930

SPECIAL ISSUE ARTICLE

WILEY

Content placement in 5G-enabled edge/core data center networks resilient to link cut attacks

Carlos Natalino[1] | **Amaro de Sousa**[2] | **Lena Wosinska**[1] | **Marija Furdek**[1]

[1]Department of Electrical Engineering, Chalmers University of Technology, Gothenburg, Sweden
[2]Instituto de Telecomuncações/DETI, Universidade de Aveiro, Aveiro, Portugal

Correspondence
Carlos Natalino, Department of Electrical Engineering, Chalmers University of Technology, Gothenburg 412 96, Sweden.
Email: carlos.natalino@chalmers.se

Funding information
This research was supported by the European Cooperation in Science and Technology, 15127. VINNOVA, SENDATE-EXTEND.

Abstract

High throughput, resilience, and low latency requirements drive the development of 5G-enabled content delivery networks (CDNs) which combine core data centers (cDCs) with edge data centers (eDCs) that cache the most popular content closer to the end users for traffic load and latency reduction. Deployed over the existing optical network infrastructure, CDNs are vulnerable to link cut attacks aimed at disrupting the overlay services. Planning a CDN to balance the stringent service requirements and increase resilience to attacks in a cost-efficient way entails solving the content placement problem (CPP) across the cDCs and eDCs. This article proposes a framework for finding Pareto-optimal solutions with minimal user-to-content distance and maximal robustness to targeted link cuts, under a defined budget. We formulate two optimization problems as integer linear programming (ILP) models. The first, denoted as K-best CPP with minimal distance (K-CPP-minD), identifies the eDC/cDC placement solutions with minimal user-to-content distance. The second performs critical link set detection to evaluate the resilience of the K-CPP-minD solutions to targeted fiber cuts. Extensive simulations verify that the eDC/cDC selection obtained by our models improves network resilience to link cut attacks without adversely affecting the user-to-content distances or the core network traffic mitigation benefits.

KEYWORDS

content caching, content delivery network, content placement, critical link set detection, link cut, malicious attacks, network resilience, optical network

1 | INTRODUCTION

Emerging applications have been pushing the limits of throughput and latency that current network deployments can offer. For instance, 4K and/or 360° video streaming, augmented and virtual reality (AR/VR), or remote machinery control applications require high throughput, low latency, and high reliability to provide satisfying user experience. Content delivery network (CDNs) are used to improve latency and robustness [17], and alleviate the traffic in core networks [14], by replicating content across large-scale data centers (DCs) at geographically disjoint locations. However, the original CDN architecture cannot offer as low latency levels as some of the emerging applications require. To support these stringent requirements in a scalable manner, the fifth generation of networks (5G) networking paradigm introduces small DCs at the network edge, where the most popular content is cached closer to the end users [24, 30].

The implementation of the models from this work are available at https://github.com/carlosnatalino/networks-5g-cdn

The introduction of edge computing has the potential to help satisfy the stringent 5G performance requirements, but poses new challenges to the infrastructure providers. For instance, the smaller scale of edge data centers (eDCs) leads to higher costs per unit of storage due to inefficiencies that arise from cooling and/or lower degree of resource sharing [2, 28]. Therefore, CDNs use a hierarchical, two-tier approach to cost-efficiently cope with the application requirements. At the upper tier, large-scale, that is, core data centers (cDCs) provide higher capacity at a lower cost per unit of storage, but users that connect to them utilize a relatively greater amount of core network resources at a higher delay. At the lower tier, small-scale, that is, eDCs serve the traffic requests locally, thus alleviating the traffic load in the core network and lowering the latency, but their use requires careful content management to achieve acceptable costs.

Optical network infrastructure is a critical enabler of CDN connectivity and is crucial to support 5G requirements at the access segment as well [27]. Conversely, inherent vulnerabilities of the optical network infrastructure can be exploited to deliberately disrupt the overlay CDN services. Attack methods which exploit optical infrastructure vulnerabilities include, for example, insertion of harmful optical signals [26] or malicious cuts of critical fiber links [20]. The effects of such malicious attacks can range from performance degradation (in terms of, e.g., throughput and latency), over cascaded congestion in the remaining network segments that take over the affected traffic, to complete disconnection of a significant portion of the network affecting hundreds of gigabits per second in the core, and substantial damage in the access segment [31].

By replicating the content at geographically distributed locations, CDNs intrinsically increase network resiliency to failures. However, targeted malicious attacks substantially differ from failures [21] because they aim at maximizing the disruption by disabling critical network components. Targeted link cut attacks that sever the most critical links are an example attack technique that is relatively easy to implement and can cause major disruption. Ignoring the threat and the effects of such attacks can lead to insufficient CDN robustness despite a high number of replicas, as shown in [17]. With the introduction of edge computing, eDCs should also be leveraged to improve the network robustness to attacks. A small portion of the most popular contents hosted at an eDC could serve a significant portion of local requests in the event of a malicious attack targeting the core network. For instance, by hosting 1% of the most popular content, an eDC could serve around 50% of the local requests for content [2, 3]. Due to the higher cost per storage unit compared to cDCs, eDCs should be carefully used to achieve the benefits that justify their higher cost. Therefore, optimizing latency and robustness while combining the use of cDCs and eDCs under cost constraints becomes very challenging.

This article proposes a framework for content placement in 5G-enabled CDNs to improve their robustness against targeted link cuts with a minimal impact on latency. In 5G-enabled CDNs, content can be reached from any cDC, which always host replicas of all contents, or from an eDC, which are sparsely deployed and host a limited amount of content replicas. Considering this scenario, two optimization problems are formulated as compact integer linear programming (ILP) models, which enable assessment of the trade-offs between the user-to-content distance and robustness to targeted link cuts by finding Pareto-optimal solutions. We first formulate the K-best content placement problem with minimal distance (K-CPP-minD), aimed at identifying the K-best content placement solutions, that is, locations and types of cDC/eDC nodes that minimize the average user-to-content distance in the network under a defined budget. Robustness of the solutions to the K-CPP-minD problem is then evaluated by the newly proposed Critical Link Set Detection (CLSD) model. CLSD finds the set of links which, if removed from the network, result in the strongest disruption, that is, disconnect the highest number of users from the content. The K-CPP-minD solution resilience to the disruption caused by cutting the p critical links identified by CLSD is evaluated in terms of the average content accessibility (ACA) [17], while μ-ACA [18] measures the mean ACA over a set of attack scenarios obtained by cutting 2 to $|p|$ links from the CLSD set.

By combining the two models, the proposed framework provides an exact method that identifies Pareto-optimal solutions in terms of average user-to-content distance and resilience to targeted link cuts. Note that the content placement problem (CPP) in a CDNs is NP-hard [4] even to find the optimal (the first best) solution. Therefore, the problem solved by our framework is strongly NP-hard as we aim to identify not only the optimal solution but also the K best solutions in the K-CPP-minD problem. This means that, in practice, there is no polynomial time algorithm able to find all Pareto-optimal solutions even for small sized problem instances. Nevertheless, the computational results will show that the proposed ILP models can be efficiently solved so that some of the Pareto-optimal solutions, the ones with smaller user-to-content distance, can be computed in reasonable time for two real-world network topologies. Extensive simulations using different eDC cost configurations verify that our approach improves network robustness without significant penalties in terms of user-to-content distance. Moreover, besides the expected user-to-content distance benefit, the introduction of eDCs reduces the traffic carried by the core network by up to 15% while maintaining or improving the CDN robustness.

The remainder of the article is organized as follows. Section 2 presents an overview of the related work. The formal problem statement and the ILP models for the K-CPP-minD and CLSD problems are described in Section 3. Section 4 presents the simulation scenario and discusses the results. Finally, Section 5 concludes the article and presents some remaining challenges.

2 | RELATED WORK

The CPP problem in CDNs has been extensively studied in the literature, with an encompassing overview given in [22]. The authors in [12] propose an optimization model for energy saving by taking into account the energy consumption of the various network segments traversed by the content. The study investigates the impact of content placement on energy consumption, and shows that the energy and the amount of generated traffic can be influenced by the tailored placement of content. In [13], the authors solve CPP jointly with allocating network resources to connections for delivering the content to users and for DC-to-DC synchronization. Results show that the inter-DC traffic can significantly impact the network load, and should be considered during content placement. The work in [30] studies the CPP variant with enabled caching at different locations in the network. The evaluation of the content access delay and the traffic load for different content caching portions and locations shows significant potential benefits of caching the content closer to the user. However, the obtained gains vary for different locations and percentage of caching. The authors in [2] study the energy-efficient CPP in metro area networks. Their strategies rely on powering cache nodes at different locations on and off to reduce energy consumption. Results show that caching at different network locations can be effective in saving energy, and that when traffic load is high, it is advisable to place the content closer to the users.

Resilience of CDNs to single-link and/or node failures is addressed in [9] by placing content replicas and assigning network resources to working and backup connections that serve user-to-DC and DC-to-DC traffic under the link- and/or DC-disjointedness constraint. CDN resilience to natural disasters is studied in [5]. ILP-based and heuristic solutions for the disaster-aware integrated CPP and connection routing is proposed in [11], where disaster zones are modeled as shared risk group (SRG). The dynamic variant of CPP is addressed in [8], where content location is defined by the experienced disaster events and the current user demand. Simulation results are obtained by assigning failure probabilities to different devices inside SRGs. The works that consider disaster-aware CPP leverage on the known correlation between network segments and devices to assess the likelihood of being disrupted by a disaster. However, malicious attacks are driven by the importance of a particular network element for proper network functioning rather than their geographical location. Therefore, this work assesses the robustness of the network considering the worst-case attack scenarios, identified by the CLSD model.

Resilience of CDNs deployed over various physical network topologies to targeted link cuts is studied in [17]. The ACA measure proposed therein captures network connectivity under anycast communication (suitable for CDNs) in the presence of such attacks. Results show that content placement plays an important role in CDN robustness to attacks. In [18], the ACA measure is extended by the μ-ACA measure, which gauges the mean robustness of a given topology and placement solution over a number of different attack scenarios. An infrastructure upgrade framework proposed therein sparsely adds links or content replicas so as to maximize the robustness, measured by μ-ACA. Link betweenness is used in both works as the criterion that guides the attack logic, but this method may lead to nonoptimal attack selection. Therefore, in this work, we formulate a CLSD model for optimal selection of links whose cutting results with maximum disruption. The problem of identifying critical elements in the network infrastructure has been the subject of several studies in the literature. Majority of works address the critical node detection problem, defined as the identification of a node set that minimizes a given connectivity metric if removed from the topology [1, 6, 23, 29]. More recently, critical link detection counterpart has also been addressed in different contexts, referring to the minimization of pairwise connectivity of communication networks [7], the minimization of the spread of infections over a population [15] and the influence propagation in social networks [10]. In this work, we formulate the CLSD problem tailored to 5G-enabled CDN which can host content in cDCs and eDC. A preliminary version of this study was presented in [16], where we investigated the trade-offs between user-to-content distance and robustness in traditional CDNs with cDCs only. In this article, we extend our previous work by considering 5G-enabled CDNs, where cDCs and eDCs can be simultaneously used to provide lower latency and improve robustness.

3 | ATTACK-RESILIENT CONTENT PLACEMENT FRAMEWORK

In this work, we consider a 5G-enabled CDN where each node in the network serves a metropolitan area, depicted in Figure 1. In this context, one of the two DC types can be colocated with any network node: cDCs and eDCs. cDCs host a replica of each content available in the CDN. Nodes colocated with a cDC handle traffic requests from the local users connected to that node, and can also serve requests from any other node in the network. eDCs host replicas of only a portion of available contents, so nodes colocated with an eDC handle only the traffic from their own local users connected through the access segment. Requests for content that is not replicated in the local eDC are forwarded to the closest cDC, contributing to core network traffic. Similarly, nodes that are not colocated with any DCs forward their requests to the closest cDC [2].

The deployment of each DC type depends on its particular properties. The cDCs are meant to handle a substantially higher number of demands than the eDCs. Due to the higher scale and efficiency of cDCs their cost per unit of processing/storage

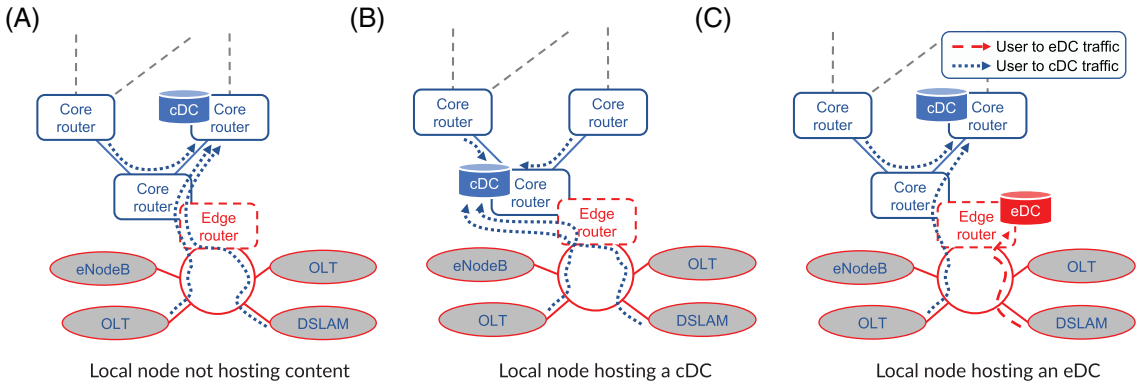

FIGURE 1 Network architecture considered in this article. Ordinary core nodes always forward their requests to the closest cDC node. Nodes hosting core data centers (cDCs) handle the traffic from their own users and from any other node. Nodes hosting edge data centers (eDCs) handle only (a part of) the traffic from their own users [Color figure can be viewed at wileyonlinelibrary.com]

tends to be lower than that of eDCs. The price difference is mainly attributed to higher energy efficiency of cDCs [2, 28] and lower operational costs. Therefore, it is expected that the cost of hosting a content replica in an eDC is substantially higher than hosting the same replica in a cDC. However, many works in the literature showed that a high percentage of traffic requests can be served by eDCs that only host a small percentage of the content [2, 3]. In this way, the eDC cost overhead can be offset by caching only a small percentage of the most popular content.

The use of eDCs also provides some unique benefits. For instance, eDC deployment can substantially reduce the communication latency (spawned by the distance and the switching devices traversed between the user and the content replica). This particular latency reduction obtained by eDCs is an important enabler of some 5G use cases. Moreover, eDCs offload the traffic from the core network, potentially reducing the total traffic carried, and allowing for a longer lifetime of currently deployed core networks [14]. The use of eDCs is also beneficial in terms of resilience. Since a part of the traffic requests from local users at nodes that host eDCs is handled locally, these demands are immune to malicious cuts of core network links and, therefore, the use of eDCs improves the robustness of the CDN to such attacks.

In the following, we provide a formal problem statement, as well as the formulation for the content replica placement and CLSD problems.

3.1 | Problem statement

When deciding on the content placement in a CDN, our goal is to decide which network nodes to host cDCs/eDCs and how to locate the content replicas across the DC nodes, under a given cost budget B. For the sake of simplicity and generality, in this work we consider unitary cost, normalized to the cost of selecting a cDC to host a replica of all contents in the network. We assume that each cDC costs 1 unit, while deploying an eDC may cost a fraction of a unit depending on how much of the content is hosted. Moreover, we consider that the cost of cDCs is the same regardless of their location. These assumptions can be easily modified in the model. The goal is to compute the content placement solutions that both (a) minimize the average user-to-content distance (which in turn minimizes the CDN latency) and (b) maximize the μ-ACA metric to maximize the CDN robustness to multiple link cut attacks.

The network topology is modeled by an undirected graph $G = (V, E)$ with a set of nodes V and a set of links E. The set of undirected links E is defined by its end nodes (i, j), where $i < j$. Any node $i \in V$ can be selected to be colocated with a cDC, which hosts a replica of all contents, or an eDC, which hosts a replica of a portion of the contents. In this way, each user request is served either by the local eDC if the requested content is replicated at this eDC, or by the closest cDC otherwise.

Possible DC configurations associated to each node are modeled with set $s = 0, 1, \ldots, S$. Configuration $s = 0$ denotes a node that is not colocated with any of the two DC types, as shown in Figure 1A. Configuration $s = S$ denotes a node that is colocated with a cDC, as shown in Figure 1B. Configurations $s = 1, \ldots, S-1$ denote a node that is colocated with an eDC, as shown in Figure 1C, where values $s = 1, \ldots, S-1$ refer to different percentages of the most popular content replicated at the eDC. To specify each configuration, the CDN operator needs to characterize the popularity of content (i.e., the percentage of requests for a particular content).

Let us assume that the content popularity is given by a distribution function (commonly used distribution is Zipf distribution where the popularity of the jth most popular item is proportional to $1/j$) [3]. Then, each configuration s is characterized by: (a) a hit ratio h_s which represents the percentage of local user requests for the content locally replicated in the eDC and (b) a cost c_{si} of using configuration s on the DC colocated with node $i \in V$. Consequently, for configuration $s = 0$ and a given node i, we

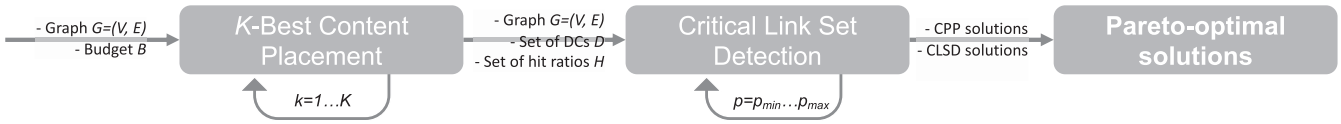

FIGURE 2 The sequence of steps used to obtain the Pareto-optimal solutions for the K-best content placement problem with minimal distance (K-CPP-minD) and critical link set detection (CLSD)

have $h_0 = 0$ and $c_{0i} = 0$; for configuration $s = S$ and a given node i, we have $h_S = 1$ (since all local users are served by the cDC), and $c_{Si} = 1$ (considering that the cost is normalized). Since higher hit ratios impose higher costs (due to higher storage, processing, and bandwidth requirements), we assume that the different configurations are defined such that $h_s < h_{s+1}$ for all s and $c_{si} < c_{s+1,i}$ for all s and i.

In general, the best content placement is not unique since the optimization problem is multiobjective. Here, we adopt the approach depicted in Figure 2 to compute the Pareto-optimal solutions. Our approach comprises two sequentially solved problems. In the first problem, the objective is to compute the k-best content placement solutions in terms of average user-to-content distance (described in Section 3.2). In the second problem, the objective is to compute the worst-case attack scenarios which minimize the μ-ACA value of each of the previous k-best content placement solutions (described in Section 3.3). Finally, by combining the solutions of both problems, it is possible to extract the Pareto-optimal solutions.

3.2 | K-CPP-minD in edge/core CDNs

The K-CPP-minD aims at computing the K-best content placement solutions in terms of the average user-to-content distance. These solutions are computed by solving a sequence of K ILP instances where: (a) the first model computes the best solution, (b) the second model computes the best solution excluding the previous one, (c) the third solution computes the best solution excluding the two previous ones, and so on. So, the kth best solution (where $1 \le k \le K$) is computed by solving the following ILP formulation which assumes that all previous solutions from 1 to $k-1$ were already been computed. In addition to the parameters introduced in Section 3.1, we also denote the shortest distance between any node pair $i, j \in V$ in graph G by δ_{ij}. The ILP formulation uses the following binary decision variables:

y_{si} defines whether or not node $i \in V$ is colocated with a cDC or an eDC; it is equal to 1 if node $i \in V$ is configured with $s = 0, \ldots, S$; and to 0 otherwise;

t_{sij} defines which cDC $j \in V$ serves requests from node $i \in V$; it is equal to 1 if the cDC at node $j \in V$ serves the user requests from node $i \in V$, and node i is configured with $s = 0, \ldots, S$; and to 0 otherwise.

In order to exclude a given CPP solution from the set of feasible solutions of an ILP formulation, we only need to know the values of the variables y_{si} of the solution. So, when computing the kth best solution, we denote the solution value of variable y_{si} of the πth solution, with $\pi = 1, \ldots, k-1$, as the binary parameter $\alpha_{si\pi}$. For a graph G and a budget B, the kth CPP solution is the optimal solution of the following K-CPP-minD(G, B) formulation (and m is the corresponding optimal value of the average user-to-content distance):

K-CPP-minD(G, B)

$$\text{Minimize} \qquad m = \left[\sum_{s=0\ldots S} \sum_{i \in V} \sum_{j \in V} \delta_{ij}(1 - h_s) t_{sij} \right] / \mid V \mid \tag{1}$$

Subject to :

$$\sum_{s=0\ldots S} \sum_{i \in V} c_{si} y_{si} \le B \tag{2}$$

$$\sum_{i \in V} y_{Si} \ge 2 \tag{3}$$

$$\sum_{s=0\ldots S} y_{si} = 1, \qquad i \in V \tag{4}$$

$$\sum_{s=0\ldots S} t_{sij} \le y_{Sj}, \qquad i \in V, j \in V \tag{5}$$

$$t_{sij} \le y_{si}, \qquad i \in V,\ j \in V,\ s = 0, \ldots, S \tag{6}$$

$$\sum_{s=0\ldots S}\sum_{j\in V} t_{sij} = 1, \qquad i \in V \tag{7}$$

$$\sum_{s=0\ldots S}\sum_{i\in V} \alpha_{si\pi} y_{si} \leq \sum_{s=0\ldots S}\sum_{i\in V} \alpha_{si\pi} - 1, \qquad \pi = 1, \ldots, k-1 \tag{8}$$

$$y_{si} \in \{0,1\}, \qquad i \in V,\ \ s = 0, \ldots, S \tag{9}$$

$$t_{sij} \in \{0,1\}, \qquad i \in V, j \in V, s = 0, \ldots, S \tag{10}$$

The objective function (1) computes the average user-to-content distance. For eDCs, it only considers the portion of traffic that is not served locally, and therefore is forwarded to the closest cDC. In this way, the user-to-content distance only accounts for the traffic that is not served locally by either a cDC or an eDC.

Constraint (2) guarantees that the content placement solution cost is within the given budget B. Constraint (3) guarantees that at least two cDCs are considered in the solution (to avoid having a single point of failure). Constraints (4) guarantee that each node $i \in V$ is used with one and only one of the possible configurations $s = 0, \ldots, S$. Constraints (5) guarantee that node $j \in V$ serving the requests from node $i \in V$ is a cDC node. Constraints (6) guarantee that the configuration index s of variable t_{sij} is consistent with the configuration used in node i. Constraints (7), together with constraints (5), guarantee that one and only one cDC node is selected to serve the requests from node $i \in V$. Constraints (8) guarantee that none of the previous CPP solutions will be considered as a solution for the current problem. Note that when using this formulation to compute the first best solution (i.e., for $k = 1$), constraints (8) are an empty set. Finally, constraints (9) and (10) are the variable domain constraints.

3.3 | CLSD in edge/core CDNs

In this work, we consider the case where malicious attacks can be launched against the CDN with the objective of disrupting services by disconnecting users from the content. To evaluate the CDN robustness to targeted link cut attacks, we define a range of scenarios that are of interest in the evaluation. This range is defined by a minimum and a maximum number of link cuts that we are interested in analyzing, denoted by p_{min} and p_{max}, respectively. For a given content placement solution obtained by solving K-CPP-minD and a number of link cuts $p : p_{min} \leq p \leq p_{max}$, it is necessary to identify the set of p links whose severing maximizes CDN disruption. This problem is commonly referred to as the CLSD problem. We use ACA to gauge the level of disruption upon cutting the p critical links identified by CLSD. Once the CLSD problem is solved for all $p = p_{min} \cdots p_{max}$, the overall CDN resilience is quantified in terms of the mean content accessibility (μ-ACA) [18].

The CLSD model requires some specific inputs, in addition to those described in Section 3.1. We denote the set of nodes adjacent to node i by V_i. V_{ij} is then defined as the set of nodes adjacent to the node with the lower degree between i and j (i.e., set V_{ij} is equal to V_i if $|V_i| \leq |V_j|$, and V_j otherwise). Recall that a K-CPP-minD solution is defined by the value of the y_{si} variables, as described in Section 3.2. Based on the variable y values, it is also possible to derive the hit ratio of the configuration used in node i, denoted here as $h_i \in H$. Moreover, it is also possible to derive the set $D = \{i \in V : y_{Si} = 1\}$ which denotes the set of nodes colocated with cDCs. Based on the set D, let us denote by F the set of ordered node pairs (i, j) such that $i \in V \setminus D$ (i.e., node i does not host a cDC) and $j \in D$ (i.e., node j hosts a cDC). Using the described notation, the ILP formulation of the CLSD problem uses the following binary variables:

- x_{ij} defines whether a link is critical or not; it is equal to 1 if link $(i, j) \in E$ is selected as a critical link; and to 0 otherwise;
- u_{ij} defines whether two nodes are connected or not after the removal of critical links; it is equal to 1 if nodes i and j, where $i < j$, are still connected when the critical link set is removed from G; and to 0 otherwise;
- v_i defines whether a node is still connected or not to a cDC after the removal of critical links; it is equal to 1 if node $i \in V \setminus D$ is still connected to at least one node in D when the critical link set is removed from G; and to 0 otherwise.

For a graph G, a number of links p and a CDN configuration defined by the set D and the hit ratio values h_i of each node $i \in V$ given as inputs, the CLSD problem (G, p, D, h_i) is defined by the following ILP model:

$$\text{CLSD}(G, p, D, H)$$

$$\text{Minimize} \qquad \text{ACA}_p = \sum_{i\in V} h_i + \sum_{i\in V\setminus D} (1 - h_i) v_i \tag{11}$$

Subject to :

$$\sum_{(i,j)\in E} x_{ij} = p \tag{12}$$

$$u_{ij} \geq 1 - x_{ij}, \qquad (i,j) \in E \tag{13}$$

$$u_{ij} \geq u_{iz} + u_{zj} - 1, \qquad i = 1,\dots,(n-1),\ \ j = (i+1),\dots,n,\ \ z \in V_{ij} \tag{14}$$

$$u_{ij} \leq v_i, \qquad (i,j) \in F \tag{15}$$

$$x_{ij} \in \{0,1\}, \qquad (i,j) \in E \tag{16}$$

$$u_{ij} \in \{0,1\}, \qquad i = 1,\dots,(n-1),\ \ j = (i+1),\dots,n \tag{17}$$

$$v_i \in \{0,1\}, \qquad i \in V \setminus D \tag{18}$$

The objective function (11) computes the portion of nodes that are still connected to a cDC after the removal of p links, in addition to the portion of traffic that is locally served by cDCs or eDCs. Constraint (12) guarantees that the set of critical links contains exactly p links. Constraints (13) guarantee that the end nodes i and j of a link $(i, j) \in E$ are connected if the link is not included in the critical link set (i.e., if $x_{ij} = 0$). Constraints (14) guarantee that any two nodes $i, j \in V : i < j$, are connected if there is one node $z \in V_{ij}$ that can communicate with both i and j. Note that, in general, we can define one constraint (14) for each node $z \neq i, j$. We minimize the number of constraints (14) by considering only nodes z adjacent to either i or j (the one with the lower degree) as defined by V_{ij}. For each node pair $(i, j) \in F$, constraints (15) set the value of variable v_i to 1 if node i, which is not a cDC node, is connected to at least one node j which is a cDC node. Finally, constraints (16)-(18) are the variable domain constraints.

The robustness of each of the K content placement solutions over multiple link cut attack scenarios is evaluated by calculating the corresponding μ-ACA. μ-ACA of a content placement solution b is given by:

$$\mu\text{-ACA}_b = \frac{1}{p_{\max} - p_{\min} + 1} \sum_{p=p_{\min}}^{p_{\max}} \text{ACA}_p, \tag{19}$$

where ACA_p is the objective value of the optimal solution of the CLSD (G, p, D, H). The μ-ACA value averages the ACA_p values over all worst-case attack scenarios for all values of p ranging from p_{min} to p_{max}.

After this evaluation, each of the K content placement solutions is characterized by its average user-to-content distance and μ-ACA value. This enables us to compute the Pareto-optimal solutions (according to Figure 2), where the set of dominated solutions is eliminated. A solution is considered as dominated if its m and μ-ACA values are both worse (or one is worse while the other is equal) than the values of at least one other solution. The remaining solutions are Pareto-optimal and represent different trade-offs between the average user-to-content distance and the resilience to link cut attacks.

4 | COMPUTATIONAL RESULTS

This section presents the computational results obtained by applying the proposed models to real-world network topologies. We first describe the setup used to solve the models and compute the evaluation metrics, followed by an investigative analysis and discussion of the obtained results.

4.1 | Setup

The results presented in this section were obtained by a custom-built Java-based tool, which writes the optimization programming language (OPL) and data files respective to the problem, and solves the problem by calling the CPLEX 12.6.3 library. All computational results were obtained on a workstation running Red Hat Enterprise Linux (RHEL) with an 8-cores 16-threads Intel Xeon processor clocked at 3 GHz and 64 GB of RAM. Each problem instance was solved optimally by CPLEX, that is, no gap was allowed, using a maximum of four parallel threads and default values for the rest of the CPLEX settings. Two publicly available core network topologies were used to obtain the results[1]: the Germany50 topology [19] with 50 nodes, 88 links, average nodal degree of 3.52, and average link length of 100 km; and the Coronet Conus topology [25] with 75 nodes, 99 links, average nodal degree of 2.64, and average link length of 329 km. The link lengths were computed using the Euclidean distance between the adjacent nodes, and considering the curvature of the Earth surface.[2]

[1]Graphical representations of the topologies are omitted for the sake of space, and can be found in our previous work [16].
[2]The page http://www.movable-type.co.uk/scripts/latlong.html describes the appropriate method to compute the distance.

TABLE 1 The considered DC node configurations, that is, core data centers (cDCs), and edge data centers (eDCs) of types 1 and 2

CDN DC type	Content hosted (%)	Hit ratio (h)	Normalized cost
cDC	100	1	1
eDC_1	1	0.5	0.1, 0.2
eDC_2	10	0.8	0.2, 0.4

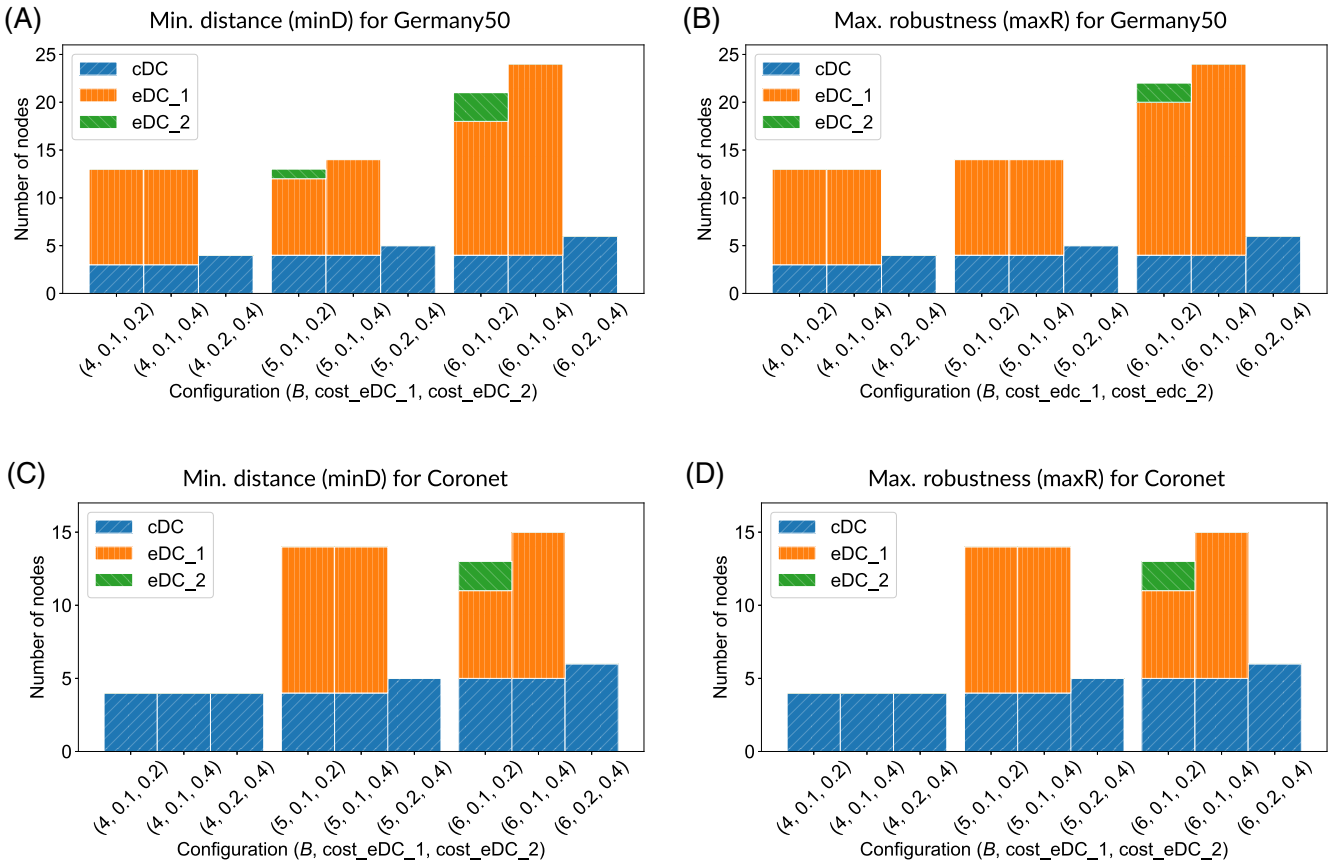

FIGURE 3 Number of nodes selected to colocate core data centers (cDCs), and edge data centers (eDCs) of type 1 and of type 2. The maxR is shown for $p_{max} = 12$ [Color figure can be viewed at wileyonlinelibrary.com]

For a given topology and values of K, B, c, p_{min}, and p_{max}, the tool solves the K-CPP-minD (described in Section 3.2) and stores, for each solution, the set of node configurations D, the average shortest-path user-to-content distance, and the set of hit ratios of each node in H. Then, for each K-CPP-minD solution, the tool solves the CLSD(G, p, D, H) model for each value of $p = p_{min}, \ldots, p_{max}$, and computes the value of μ-ACA (described in Section 3.3).

The use of eDCs is modeled by a set of possible configurations representing different percentages of the most popular content replicated at the eDCs. We assume that CDN users are attached to all nodes in the network and the aggregate user request rate is similar for all nodes. We consider the content popularity given by a Zipf distribution (i.e., the popularity of the ith most popular item is proportional to $1/i$) [2, 3]. Table 1 summarizes the considered configuration. Each cDC hosts all contents available in the CDN, and costs 1 unit ($c_{Si} = 1$, with $S = 3$). To avoid the problem of having a single point of failure, we consider that at least two cDCs are placed in the network. The eDC of type 1 replicates 1% of the content, but serves 50% of all the requests of contents from that node. The eDC of type 2 replicates 10% of the content, which serves 80% of all the requests of contents from that node. For each eDC type, we consider two different cost values, resulting with a total of 3 different cost configurations, that is, ($c_{1i} = 0.1$, $c_{2i} = 0.2$); ($c_{1i} = 0.1$, $c_{2i} = 0.4$); and ($c_{1i} = 0.2$, $c_{2i} = 0.4$). This range of configurations allows us to also analyze the trade-offs in terms of cost.

For both considered topologies, we solve the K-CPP-minD problem for three different budget values, that is, $B = 4$, 5 and 6, and compute the $K = 2000$ best solutions. In all cases, μ-ACA is computed considering $p_{min} = 2$ ($p_{min} = 1$ is not applicable as all topologies are 2-connected), while p_{max} values of 6, 9 and 12 are considered in order to assess a wide range of disruption scenarios.

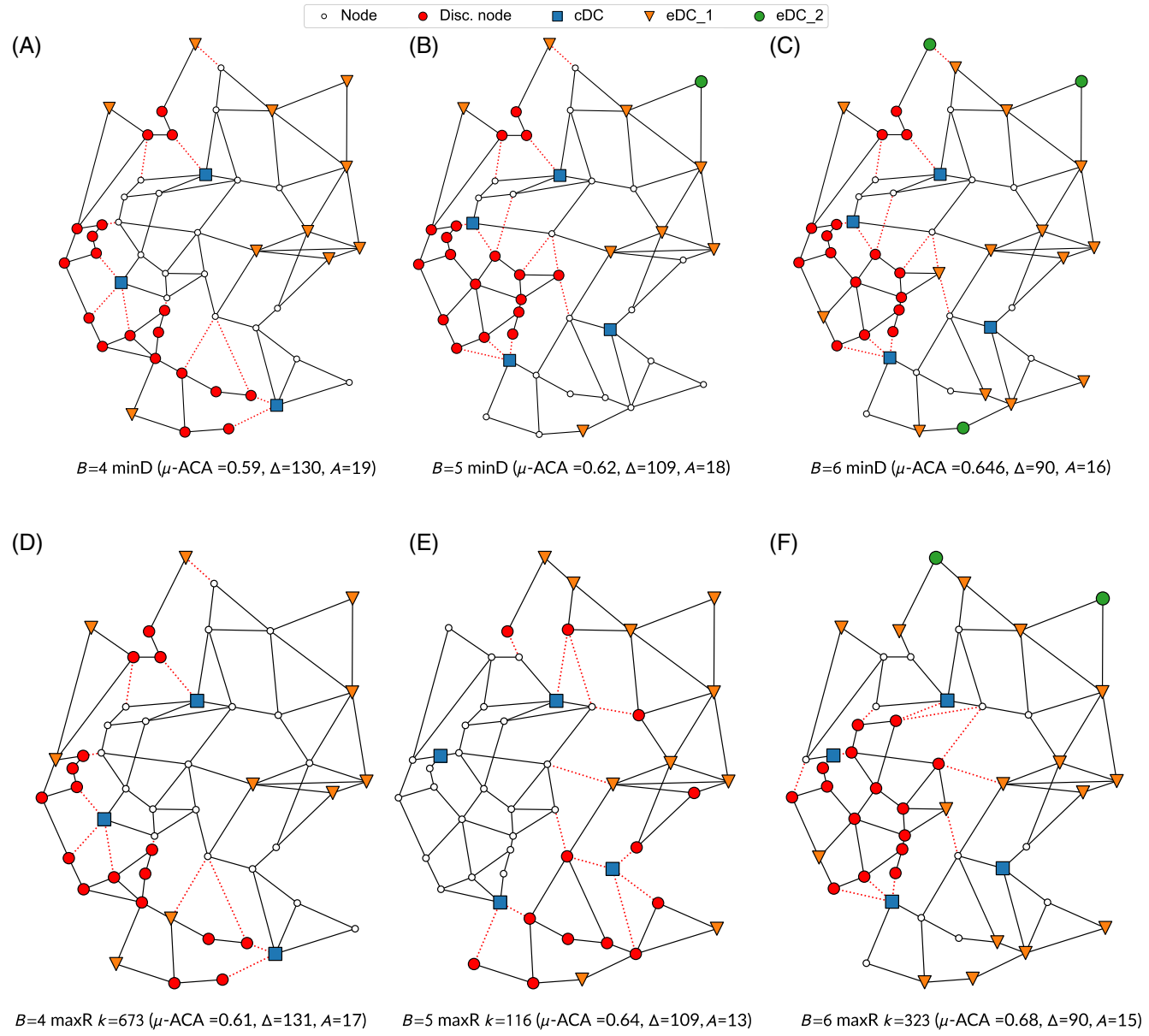

FIGURE 4 Content placement solutions with different DC types (cDCs and eDCs of type 1 and of type 2) for $c = (0.1, 0.2)$, critical links (red dotted lines) and disconnected (red larger) nodes of the first solution (minD) and maximum robustness solution (maxR). All solutions are shown for $p = 12$. The maxR is shown for $p_{max} = 12$. Δ represents the average user-to-content distance, while A represents the number of disconnected nodes [Color figure can be viewed at wileyonlinelibrary.com]

Among all solutions for each particular configuration, we compute the Pareto-optimal solutions, that is, the solutions that represent the different trade-offs between the average user-to-content distance and the attack resilience in terms of μ-ACA. Moreover, for each configuration, we record the solution with minimum distance (denoted as *minD*), which is always the first K-CPP-minD solution, and the solution with maximum robustness (denoted as *maxR*), which yields the highest μ-ACA among all K-CPP-minD solutions.

In addition to the average user-to-content distance and robustness evaluation, we also evaluate the traffic carried by the core network in each scenario. In this case, we are particularly interested in the amount of network resources necessary to serve the total CDN traffic in the core network. We generalize this concept by computing the outgoing traffic from each node, and the number of links that this traffic traverses. All traffic flowing out of a node always connects to the closest cDC. The amount of traffic that flows out of each node is computed as follows. A node that hosts a cDC will not have any outgoing traffic; a node that hosts an eDC of type 1 will have 50% of its traffic served locally and 50% served by the closest cDC; an eDC of type 2 will serve 80% of its traffic and 20% will be served by the closest cDC. Finally, we summarize the number of links traversed by each flow, weighted by the percentage of the node traffic carried by the flow. To provide a more general analysis, we normalize the total traffic carried by the traffic in a cDC-only network (no eDCs are available).

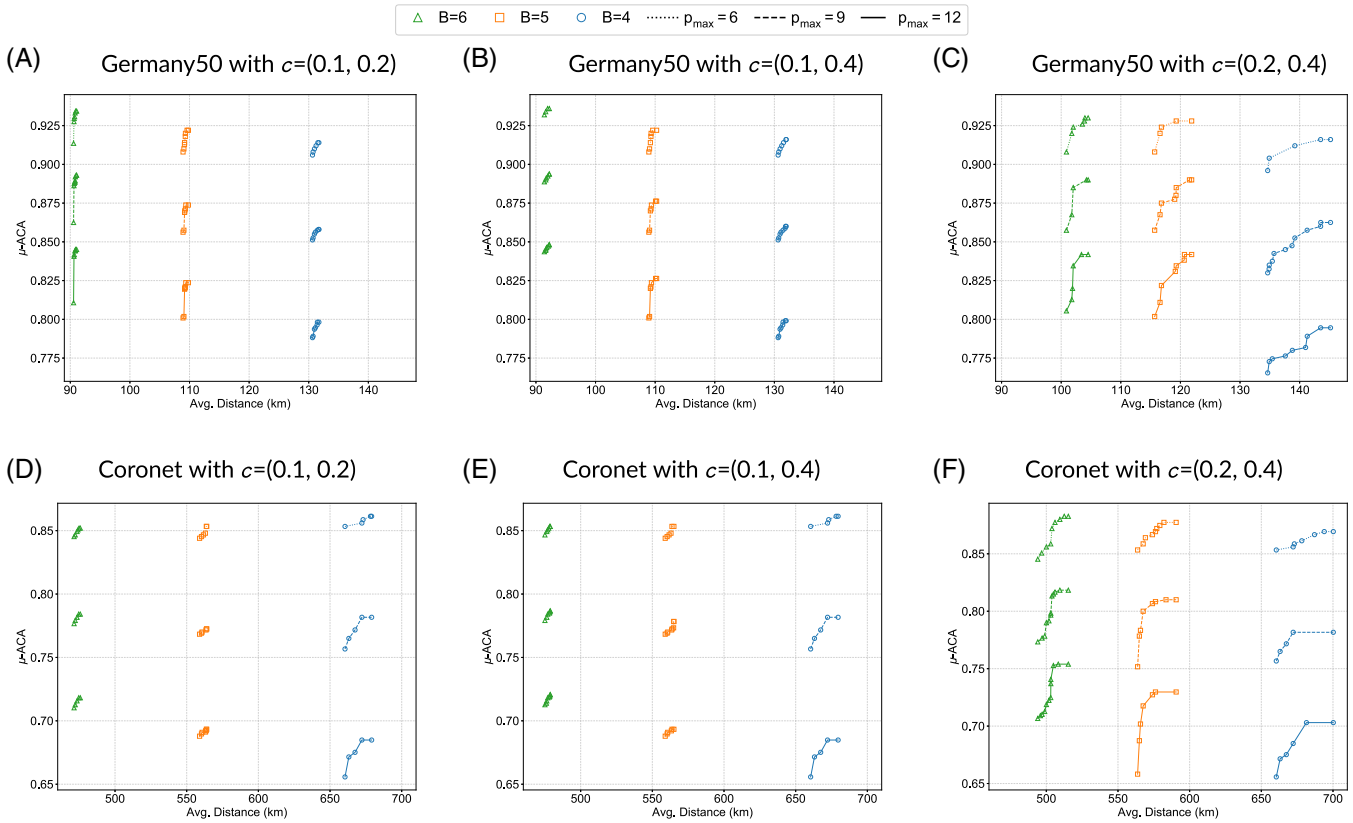

FIGURE 5 Average shortest user-to-content distance vs μ-ACA of the Pareto-optimal solutions for different eDC cost configurations (c). Line styles identify the maximum number of critical links (p_{max}) while markers identify the budget (B) [Color figure can be viewed at wileyonlinelibrary.com]

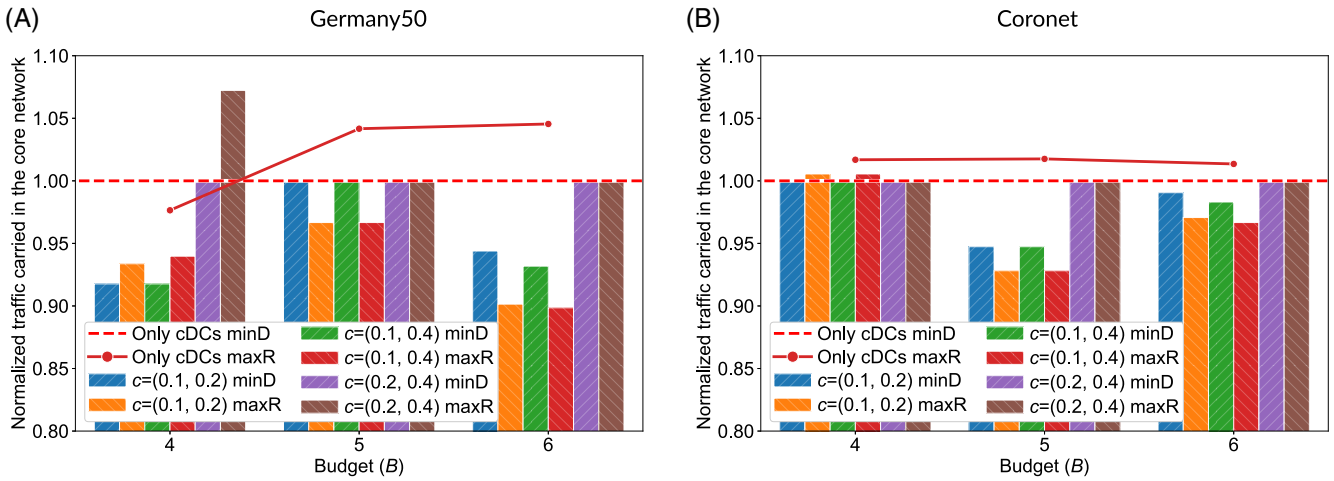

FIGURE 6 Traffic carried by the core network when using the minimum distance solution (minD) and the maximum robustness solution (maxR) computed for $p_{max} = 12$, normalized by the traffic obtained when only cDCs are used, for different eDC cost configurations (c) [Color figure can be viewed at wileyonlinelibrary.com]

4.2 | Performance assessment

Figure 3 presents the number of nodes selected as cDCs and eDCs for different values of budget B and eDC cost configurations in the two topologies. When the cost of eDCs is the highest (i.e., 0.2 and 0.4), eDCs are not selected in any scenario. This shows that, at the highest cost considered, eDCs are too expensive, and present no benefits in terms of user-to-content distance or robustness. When eDC costs are lower, for Germany50 network (Figures 3A,B), the more robust solution trades eDCs of type 2 for eDCs of type 1, compared to the minD solution. This shows a general trend in the solutions, where eDCs of type 1 are used more often than eDCs of type 2 due to the lower cost (allowing their deployment at a larger number of nodes for the same budget), but fairly high hit ratio. For the Coronet Conus topology (Figure 3C,D), eDCs are not used in solutions with the lowest budget ($B = 4$), and in other cases, the number of nodes of each DC type does not change. In general, this shows that

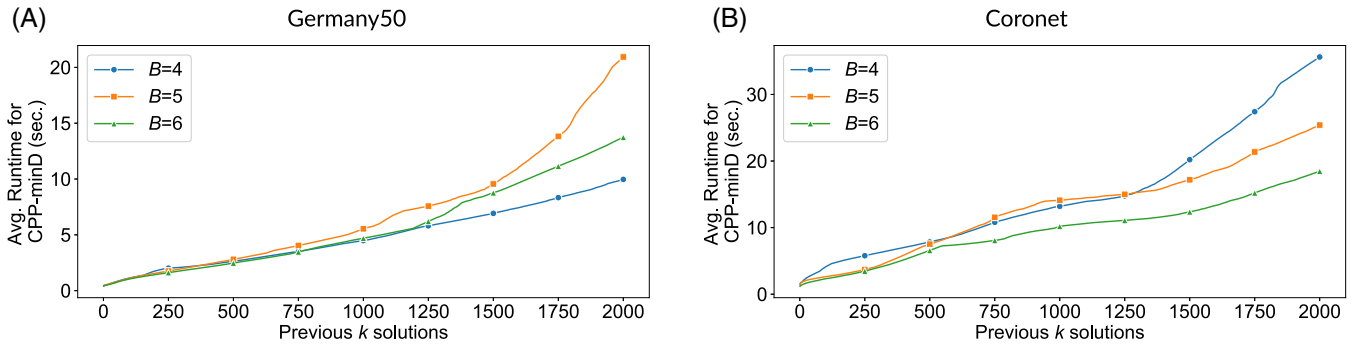

FIGURE 7 Average CPLEX runtime to solve the K-CPP-minD problem for different budgets (B) as a function of the previous k-best replica placement solutions [Color figure can be viewed at wileyonlinelibrary.com]

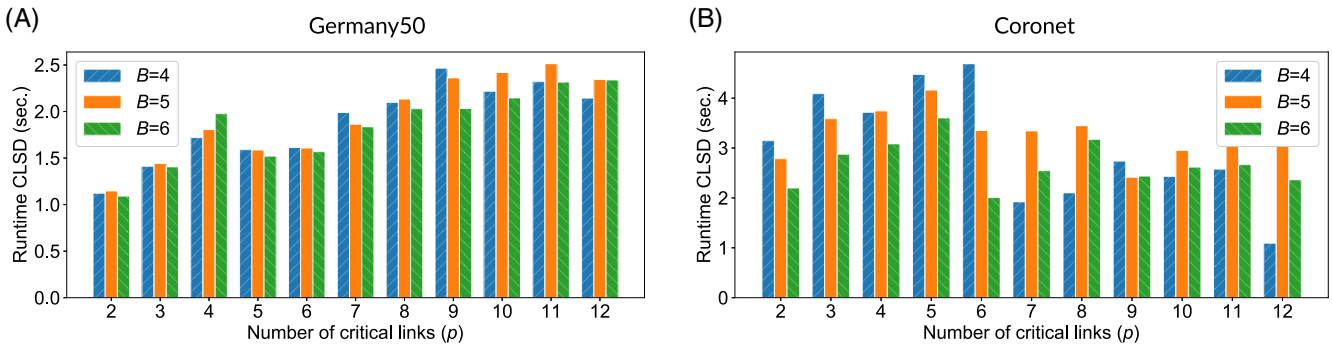

FIGURE 8 Average CPLEX runtime to solve the K-CPP-minD problem for different budgets (B) as a function of the previous k-best replica placement solutions [Color figure can be viewed at wileyonlinelibrary.com]

the placement of DC nodes makes the biggest difference for robustness, and adjusting the proportion of the different DCs helps improving robustness in some cases.

Figure 4 shows the solution for Germany50 topology with two representative cases for $B = 4$, 5, and 6, when the cost of eDCs is set to the lowest considered value, that is, (0.1, 0.2). The solutions shown in the top row are the ones with the minimal user-to-content distance (minD), that is, the first solution of the K-CPP-minD. The solutions in the bottom row are the ones with the highest resilience to link cut attacks (maxR). Note that, for each budget, the maxR solution is found at a different k. As expected from Figure 3A,B, the number of nodes hosting each DC type changes only when $B = 5$. In other cases, the difference between minD and maxR solutions is observed in the placement of the nodes. To obtain higher robustness, the main changes in DC positioning are observed for eDCs of type 1, which are placed in regions impacted by the link cuts to support nodes disconnected from the cDCs. Moreover, with the increase in budget from $B = 5$ to $B = 6$, the higher budget is entirely invested in eDCs, showing that their use in CDNs is beneficial both in terms of user-to-content distance and resilience.

Figure 5 shows the Pareto-optimal solutions for all test cases of the Germany50 and Coronet Conus topologies under different cost configurations (c) and budgets (B). As expected, the μ-ACA values decrease with the increase of p_{max}, due to the larger disruption caused by a higher p_{max}. Moreover, in almost all cases, a significant improvement in μ-ACA can be achieved at the expense of a small increase in the average user-to-content distance. For instance, in Figure 5A, for $B = 6$, there is a substantial improvement in μ-ACA when the average user-to-content distance increases by only a few meters. In the Coronet Conus network with $B = 4$, when eDCs have the highest cost, that is, $c = (0.2, 0.4)$, more points in the Pareto-optimal solutions are observable. Recall that, as shown in Figure 3, in these cases only cDCs are selected, and average user-to-content distance is higher than for the cases where eDCs are selected.

Figure 6 shows the normalized CDN traffic carried by the core network for the two topologies. In this case, a value lower than one means that the amount of traffic entering the core reduces, thus alleviating the core network for other types of traffic, or delaying the end of life of the technology used. In general, we can observe that when only cDCs are available, the solutions with the highest μ-ACA (maxR) usually result with more traffic flowing through the network (except for the Germany50 topology with $B = 4$). On the contrary, when eDCs are available, solutions with the highest robustness (maxR) also reduce the amount of traffic flow (again, except the particular case of Germany50 topology with $B = 4$). These trends show that the use of eDCs does not only help improve network resilience to targeted fiber cuts, but also alleviates the traffic overhead in the core network. The Germany50 topology benefits the most with budgets equal to 4 or 6, as shown in Figure 6A. The Germany50 maxR solution under $B = 6$, reduces the traffic flowing through the core network by up to 10% compared to the cDC-only minD case, and up

to 15% compared to the cDC-only maxR case. For the Coronet Conus topology with $B = 5$, the traffic can be reduced by more than 10%, also observed for configurations with the highest robustness.

The average running times to solve the K-CPP-minD and CLSD ILPs for the Germany50 topology are shown in Figures 7A and 8A. Figure 7A shows how the running time for the K-CPP-minD problem increases with the number of previous solutions (k). This behavior is expected since the higher is k, the more constraints need to be added to the model (as defined in (8)). Moreover, the budget also impacts the runtime for higher values of k, since the number of nodes that are selected to host replicas are higher. For the CLSD problem, Figure 8A shows that the average runtime is below 3 seconds which indicates compactness of the formulation.

The higher number of nodes and links in Coronet Conus topology than in Germany50 reflects on the runtimes presented in Figures 7B and 8B. In general terms, the runtime for Coronet Conus is around twice higher than for Germany50. However, the runtimes for Coronet Conus present some different trends. In Figure 7, while the K-CPP-minD highest runtime for Germany50 is observed with $B = 5$, for Coronet Conus the trends is descending with the budget, with $B = 4$ taking the longest, and $B = 6$ taking the shortest time, on average. In Figure 8, while the CLSD runtime increases with p for Germany50, this trend cannot be observed for Coronet Conus, where higher p often presents lower average runtime.

5 | CONCLUSIONS

This article addressed the challenge of robust placement of content in 5G-enabled CDNs that can replicate content at cDC and eDCs. The problem was formulated as two optimization models that minimize the average user-to-content distance and maximize the robustness in a coordinated way. The proposed framework leveraged these two models to identify the Pareto-optimal solutions, showing the trade-offs between the two different objectives, while respecting a defined budget.

Simulation results obtained by applying the proposed framework on two real-world network topologies showed that the network resilience to targeted link cuts can be significantly improved at the expense of a small increase in the average user-to-content distance. Moreover, the assessment of the traffic traversing the core network showed that eDCs also reduce the load in the core network, potentially providing several benefits. In all cases, the runtime needed to solve the models to their optimality remained very low even for the medium-high size topologies considered.

For future work, studying an integrated optimization model that can jointly minimize user-to-content distance and maximize robustness would be interesting. Moreover, assessing the impact of attacks to the metro/access optical networks in metropolitan areas would provide a fine-grained assessment of the impact of targeted attacks in users. Analyzing the robustness of wireless/wired fronthaul/backhaul networks considering edge computing is another interesting problem to be addressed in the future.

ORCID

Carlos Natalino https://orcid.org/0000-0001-7501-5547
Amaro de Sousa https://orcid.org/0000-0002-5804-1337
Lena Wosinska https://orcid.org/0000-0001-6704-6554
Marija Furdek https://orcid.org/0000-0001-5600-3700

REFERENCES

[1] A. Arulselvan, C.W. Commander, L. Elefteriadou, and P.M. Pardalos, *Detecting critical nodes in sparse graphs*, Comput. Oper. Res. **36** (2009), 2193–2200.
[2] O. Ayoub, F. Musumeci, M. Tornatore, and A. Pattavina, *Energy-efficient video-on-demand content caching and distribution in metro area networks*, IEEE Trans. Green Commun. Netw. **3** (2019), 1–1.
[3] E. Bastug, M. Bennis, and M. Debbah, *Living on the edge: The role of proactive caching in 5G wireless networks*, IEEE Commun. Mag. **52** (2014), 82–89.
[4] F. Chen, K. Guo, J. Lin, and T. La Porta, "Intra-cloud lightning: Building CDNs in the cloud," Proceedings IEEE INFOCOM, Orlando, FL, 2012, pp. 433–441.
[5] C. Colman-Meixner, C. Develder, M. Tornatore, and B. Mukherjee, *A survey on resiliency techniques in cloud computing infrastructures and applications*, IEEE Commun. Surv. Tutor. **18** (2016), 2244–2281.
[6] M. Di Summa, A. Grosso, and M. Locatelli, *Branch and cut algorithms for detecting critical nodes in undirected graphs*, Comput. Optim. Appl. **53** (2012), 649–680.
[7] T.N. Dinh, Y. Xuan, M.T. Thai, P.M. Pardalos, and T. Znati, *On new approaches of assessing network vulnerability: Hardness and approximation*, IEEE/ACM Trans. Netw. **20** (2012), 609–619.
[8] S. Ferdousi, F. Dikbiyik, M.F. Habib, M. Tornatore, and B. Mukherjee, *Disaster-aware datacenter placement and dynamic content management in cloud networks*, IEEE/OSA J. Opt. Commun. Netw. **7** (2015), 681–694.

[9] M. Furdek, A. Muhammad, and L. Wosinska, "Survivable manycast, anycast and replica placement in optical inter-datacenter networks," 19th International Conference on Transparent Optical Networks (ICTON), Girona, 2017, pp. 1–4.
[10] C.P. Gillen, A. Veremyev, O.A. Prokopyev, and E.L. Pasiliao, *Critical arcs detection in influence networks*, Networks **71** (2017), 412–431.
[11] M.F. Habib, M. Tornatore, M. De Leenheer, F. Dikbiyik, and B. Mukherjee, *Design of disaster-resilient optical datacenter networks*, J. Lightw. Technol. **30** (2012), 2563–2573.
[12] R. Modrzejewski, L. Chiaraviglio, I. Tahiri, F. Giroire, E. Le Rouzic, E. Bonetto, F. Musumeci, R. Gonzalez, and C. Guerrero, "Energy efficient content distribution in an ISP network," IEEE Global Communications Conference (GLOBECOM), Atlanta, GA, 2013, pp. 2859–2865.
[13] A. Muhammad, N. Skorin-Kapov, and M. Furdek, *Manycast, anycast and replica placement (MARP) in optical inter-datacenter networks*, IEEE/OSA J. Opt. Commun. Netw. **9** (2017), 1161–1171.
[14] F.J.M. Muro, N. Skorin-Kapov, and P. Pavon-Marino, *Revisiting core traffic growth in the presence of expanding CDNs*, Comput. Netw. **154** (2019), 1–11.
[15] A.K. Nandi and H.R. Medal, *Methods for removing links in a network to minimize the spread of infections*, Comput. Oper. Res. **69** (2016), 10–24.
[16] C. Natalino, A. de Sousa, L. Wosinska, and M. Furdek, "On the trade-offs between user-to-replica distance and CDN robustness to link cut attacks," 10th International Workshop on Resilient Networks Design and Modeling (RNDM), Longyearbyen, 2018, pp. 1–7.
[17] C. Natalino, A. Yayimli, L. Wosinska, and M. Furdek, "Content accessibility in optical cloud networks under targeted link cuts," International Conference on Optical Network Design and Modeling (ONDM), Budapest, 2017, pp. 1–6.
[18] C. Natalino, A. Yayimli, L. Wosinska, and M. Furdek, *Infrastructure upgrade framework for content delivery networks robust to targeted attacks*, Opt. Switch. Netw. **31** (2019), 202–210.
[19] S. Orlowski, R. Wessäly, M. Pióro, and A. Tomaszewski, *SNDlib 1.0—Survivable network design library*, Networks **55** (2010), 276–286.
[20] J. Rak, D. Hutchison, E. Calle, T. Gomes, M. Gunkel, P. Smith, J. Tapolcai, S. Verbrugge, and L. Wosinska, "RECODIS: Resilient communication services protecting end-user applications from disaster-based failures," 18th International Conference on Transparent Optical Networks (ICTON), Trento, 2016, pp. 1–4.
[21] D.F. Rueda, E. Calle, and J.L. Marzo, *Robustness comparison of 15 real telecommunication networks: Structural and centrality measurements*, J. Netw. Syst. Manag. **25** (2017), 269–289.
[22] M.A. Salahuddin, J. Sahoo, R. Glitho, H. Elbiaze, and W. Ajib, *A survey on content placement algorithms for cloud-based content delivery networks*, IEEE Access **6** (2018), 91–114.
[23] D. Santos, A. de Sousa, and P. Monteiro, *Compact models for critical node detection in telecommunication networks*, Electron. Notes Discrete Math. **64** (2018), 325–334 8th International Network Optimization Conference—INOC 2017.
[24] W. Shi, J. Cao, Q. Zhang, Y. Li, and L. Xu, *Edge computing: Vision and challenges*, IEEE Internet Things J **3** (2016), 637–646.
[25] J. Simmons, *Optical Network Design and Planning*, 2nd edn, Springer, Switzerland, 2014.
[26] N. Skorin-Kapov, M. Furdek, S. Zsigmond, and L. Wosinska, *Physical-layer security in evolving optical networks*, IEEE Commun. Mag. **54** (2016), 110–117.
[27] B. Skubic, M. Fiorani, S. Tombaz, A. Furuskär, J. Mårtensson, and P. Monti, *Optical transport solutions for 5G fixed wireless access [invited]*, IEEE/OSA J. Opt. Commun. Netw. **9** (2017), D10–D18.
[28] A. Tzanakaki, M. Anastasopoulos, I. Berberana, D. Syrivelis, P. Flegkas, T. Korakis, D.C. Mur, I. Demirkol, J. Gutiérrez, E. Grass, Q. Wei, E. Pateromichelakis, N. Vucic, A. Fehske, M. Grieger, M. Eiselt, J. Bartelt, G. Fettweis, G. Lyberopoulos, E. Theodoropoulou, and D. Simeonidou, *Wireless-optical network convergence: Enabling the 5G architecture to support operational and end-user services*, IEEE Commun. Mag. **55** (2017), 184–192.
[29] A. Veremyev, V. Boginski, and E.L. Pasiliao, *Exact identification of critical nodes in sparse networks via new compact formulations*, Optim. Lett. **8** (2014), 1245–1259.
[30] X. Wang, M. Chen, T. Taleb, A. Ksentini, and V.C.M. Leung, *Cache in the air: Exploiting content caching and delivery techniques for 5G systems*, IEEE Commun. Mag. **52** (2014), 131–139.
[31] E. Wong, E. Grigoreva, L. Wosinska, and C.M. Machuca, *Enhancing the survivability and power savings of 5G transport networks based on DWDM rings*, IEEE/OSA J. Opt. Commun. Netw. **9** (2017), D74–D85.

How to cite this article: Natalino C, de Sousa A, Wosinska L, Furdek M. Content placement in 5G-enabled edge/core data center networks resilient to link cut attacks. *Networks.* 2020;75:392–404. https://doi.org/10.1002/net.21930